DOCUMENT RESUME

ED 460 147 TM 033 621

AUTHOR Sultana, Qaisar
TITLE The University Writing Requirement: A Study of the Reliability of Scores.
PUB DATE 2001-11-00
NOTE 11p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (30th, Little Rock, AR, November 14-16, 2001).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Students; Correlation; Essays; Higher Education; *Reliability; *Scores; Scoring; *Writing Tests
IDENTIFIERS Eastern Kentucky University

ABSTRACT

This study examined the reliability of scores assigned to the essays written by Kentucky students to meet the University Writing Requirement (UWR) at Eastern Kentucky University. Two sets of essays, 50 each, on the same prompt that had been read and scored in 1989 and 1997 by trained UWR scorers were read by 7 UWR scorers in 2000. A correlation was performed between the initial scores and the scores assigned in 2000. Correlation coefficients of 0.49 and 0.78 respectively were found between the 1989-2000 and 1997-2000 scores. Both correlations were positive. The correlation between 1989-2000 scores was weaker than the correlation between 1997-2000 scores. It appears that the reliability of scores among the UWR scorers was weaker over an 11-year period (1989-2000) compared to a 3-year period (1997-2000). It is concluded that the reliability of scores among UWR readers was going down with the passage of time. UWR policy, scoring rubric, and writing prompt statement are appended. (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 460 147

The University Writing Requirement - A Study of the Reliability of Scores

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

By

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Q Sultana
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

Qaisar Sultana
Eastern Kentucky University

Presented

at

The 30th annual conference of Mid-South Educational Research Association

Little Rock, AR.

November 14-16, 2001

TM033621

BEST COPY AVAILABLE

ABSTRACT

This study examined the reliability of scores assigned to the essays written by Kentucky students admitted at this University to meet the University Writing Requirement (UWR). Two sets of essays, 50 each, on the same prompt, read and scored in 1989 and 1997 by trained UWR scorers were read by seven UWR scorers in year 2000. A correlation was performed between the initial scores and the scores assigned in 2000. A correlation coefficient of .49 and .78 respectively was found between the 1989-2000 and 1997-2000 scores. Both correlations were positive. The correlation between 1989-2000 was weaker than the correlation between 1997-2000 scores. It appeared that the reliability of scores among the UWR scorers was weaker over an eleven-year period (1989-2000) compared to a three-year period (1997-2000). It was concluded that the reliability of scores among UWR readers was going down with passage of time.

The publication of *A Nation at Risk*, in 1983 drew national attention to the quality of education in the United States. This publication gave impetus to various initiatives aimed at improving education in the nation. The country embarked on educational reform. Universities and colleges took stock of their own education programs. The Carnegie Report and the Holmes Group came up with their recommendations to improve teacher education. Almost every state in the nation enacted its own initiatives to reform education.

At this University, a small group of faculty members, concerned about the poor communication skills of the students, particularly in written language, formed a committee. After a few years of deliberations, this group, called the University Writing Requirement Committee, proposed a University Writing Requirement (UWR) for all EKU students seeking a baccalaureate degree. Appropriate authorities of the university approved this proposal. It was implemented in 1989. Since then, students are required to take the UWR in the first semester following completion of 60 credit hours of course work. Transfer students who transfer 60 credit hours or more must take the exam in the first semester of enrollment. Students who fail the first attempt may retake the exam and continue taking courses in the following semester(s) under certain conditions. (A complete copy of the UWR policy is contained in Table1).

Insert Table 1 about here

A faculty member trained in holistic scoring trained a group of faculty members to score the students' essays according to a rubric, presented in Table 2, developed by the UWR Committee. The trainer continues to recruit and train more UWR readers every year or two as the need arises. The UWR is administered each semester and in the summer.

Insert Table 2 about here

The UWR consists of an essay written by the students on a prompt, which is created or selected by the UWR Committee. Occasionally a prompt is repeated. Students are given a four-page booklet with the prompt already printed on it. They are to identify themselves on the booklets only by their social security numbers. Students have an hour to write the essay. Limited English proficiency (LEP) students are allowed an additional hour. All students are permitted to bring and consult a dictionary and or a thesaurus.

In the following week, UWR readers have a scoring session. The UWR Coordinator distributes a few essays, already scored, as benchmarks. The UWR scorers read the essays. A brief explanation is provided explaining the assigned scores. This is followed by a short warm-up session conducted by the UWR Coordinator to achieve consistency among scorers. After the warm-up session, each scorer is given a stack of essays to read, assign scores according to the seven-point rubric, and enter them in the designated square on the essay booklets. On each booklet, scorers also record a number assigned to them for identification. A secretary collects the booklets as they are read and blocks the scores assigned by the first reader. The booklets are then given to a second reader who reads and records the scores on them. Thus, each essay is scored by at least two readers and occasionally by a third reader when the scores given by the first two readers are not contiguous. Two readers' scores are added and are recorded on the

booklets as the final scores. When an essay is read by three readers, the average of the scores is computed and multiplied by two. The total is recorded as the final score.

The deliberations of UWR Committee and the implementation of UWR at EKU coincided with a major event in the Commonwealth of Kentucky. In November 1985, a complaint was filed by 11 school districts in the Franklin Circuit Court in Kentucky challenging the equity and adequacy of funds provided to individual school districts by the Commonwealth. The "Circuit Court issued a judgment in October, 1988, stating that the General Assembly had failed to provide an efficient system of common schools, and that the system of school financing was inefficient, in the constitutional sense, and discriminatory." (KDE, 1994)

"On appeal, the Kentucky Supreme Court issued an opinion in June, 1989, which held that the system of common schools in Kentucky was unconstitutional." (KDE, 1994) In response to the ruling of the highest court in the Commonwealth, the General Assembly embarked on restructuring the entire public education system in Kentucky. It appointed a Task Force on Education Reform in July,1989, composed of the leadership of the House and Senate and appointees of the governor. The recommendations of the Task Force resulted in House Bill 940, which was approved by the 1990 General Assembly. The governor signed the Kentucky Education Reform Act (KERA) on April 11, 1990.

One of the key provisions of KERA is accountability. A testing program, Kentucky Instructional Retrieval Information System (KIRIS), made up of subject matter tests, performance events, and portfolios (including the writing portfolio) were administered to students in the fourth, eight, and twelfth grades during 1991-1992 school year to determine baseline data. "Portfolios occupied a key place in KIRIS, both as a means of assessment that directly tapped student work in classrooms, schools, and districts. Since the contents of the portfolios arose from student's classroom work, the portfolio was the assessment component that most clearly reflected local curriculum and instruction." (KDE, 2000) The statewide testing program, now known as the Commonwealth Accountability Testing System (CATS), continues to be administered each year and is used for rewards and sanctions awarded to schools or levied against schools across the commonwealth. The Writing Portfolio is now required in grades 4, 7, and 12.

The Writing Portfolio, designed by a committee of Kentucky English/Language Arts educators, consists of a collection of students' written products in broad categories:

- Personal experience writing;
- Imaginative writing;
- Reflective writing;
- Trans-active writing for real-world purposes and audiences.

The Writing Portfolios are scored locally by school teachers who have been provided extensive training in portfolio scoring. Six criteria are applied holistically to produce a single final judgment, Novice, Apprentice, Proficient, or Distinguished. The six criteria are:

- Purpose/Audience Awareness
- Idea development/Support;

- Organization;
- Sentence Structure and Variety;
- Language (Word Choice and Usage); and,
- Correctness (spelling, punctuation, and capitalization).

The statewide assessment system in Kentucky is designed to measure the schools' success in the attainment of Kentucky's Academic Expectations. One of the Academic Expectations is that all students should write for multiple purposes in multiple forms for a variety of audiences. The Writing Portfolio is calculated to assess this Expectation.

Last year, this researcher conducted a study comparing the writing skills, as measured by UWR, of Kentucky high school graduates who had gone through CATS writing portfolio (post-KERA) with their peers prior to the implementation of KERA. Specifically, the objectives were to (a) compare the UWR scores of pre-and post KERA students, and (b) determine the significance of the difference. A total of 50 UWR essays written by Kentucky students in 1989 (pre-KERA) and 50 UWR essays written by Kentucky students on the same prompt in 1997, both randomly selected, were used for the study. Both sets of essays were read and scored by seven UWR readers just in the same manner in which they are scored at real UWR scoring sessions. Six of the seven readers had come on board in the1990s and could not have read/scored the essays in 989. The null hypothesis was that there was no difference between the mean UWR scores of the two groups of students. The results of this study showed no statistically significant difference between the pre-and post KERA students' writing skills. These results were unexpected and prompted the researcher to engage in this study.

Purpose

The purpose of this study was to examine the consistency of scoring among the UWR scorers over the years. Since 1989, the original trainer as well as her successor have retired. Many of the original readers have either retired or have discontinued scoring the UWR essays. The scoring rubric received a minor cosmetic revision in 1992 and has remained unchanged since. This researcher wanted to investigate the reliability of UWR scorers over the years to help further analyze the results of last year's research.

Methodology

The investigator noted the scores originally received by 50 Kentucky high school graduates in 1989 (pre-KERA) and the scores assigned to the same group of essays by readers in year 2000. The mean and standard deviation were computed for each group of scores. The two scores, given in 1989 and in 2000, were compared which resulted in a Pearson correlation of 0.49. The prompt used in 1989 was also used in 1997, presented in Table 3. The investigator, therefore, randomly selected 50 essays written by Kentucky high school graduates from this group (post-KERA). The procedure followed for the pre-KERA group was performed on this group as well.

Insert Table 3 about here

This comparison between the scores assigned to this group in 1997 and those assigned in 2000 resulted in a correlation of 0.78. Complete statistics are reported in Table 4.

Insert Table 4 about here

Results

The mean score of the essays scored by UWR readers in 1989 was 8.18. The same essays scored by readers in 2000 yielded a mean score of 8.42. The 2000 mean score was slightly higher than the original mean score given in 1989. The difference between the two means was not statistically significant. The correlation coefficient between the UWR readers of 1989 and 2000 was 0.49.

The mean score of the essays scored by UWR readers in 1997 was 8.16. According to the scores assigned to the same group of essays by UWR readers in 2000, the mean was 8.20. Again, the difference between the two mean was not statistically significant. And, again the mean score according to the scores assigned in 2000 was slightly higher than the mean of the scores originally assigned to the same essays in 1997. The correlation coefficient between the two scores was 0.78. Both correlations were positive.

Discussion

The results of this study show that although the UWR scorers in year 2000 used the same rubric that had been used in the past, the interpretation and application of the rubric had become somewhat lax or lenient. The mean of the scores assigned to the essays in year 2000 is slightly higher than the one assigned to the same essays in1989 and 1997 when they were initially scored. However, the difference between the means is not statistically significant. The correlation coefficient between the1989-2000 is lower, than the correlation between the 1997-2000 group, 0.49 and 0.78 respectively. Both correlations are positive and significant at the .05 level. The results show that with increased passage of time the inter-reader reliability of UWR scorers is getting lower. A slight inflation in UWR scoring also appears to be occurring with passage of time.

The slight inflation in scores may be attributed to the change in UWR leadership. The initial UWR Coordinator, a professor in the English department, retired in 1989. An experienced UWR reader from the department of Mass Communication assumed the responsibilities of UWR Coordinator. The UWR Coordinator recruits and trains new UWR readers in scoring. The Coordinator also conducts the warm-up training at each scoring session.
A Coordinator coming from a different discipline may have contributed to a slightly different interpretation of the scoring rubric.

A cutoff score to pass the UWR is somewhat fluid. It is tentatively set at a total score of 7. However if the number of students scoring below the standard, a total score of 7, is more than a acceptable percentage of the total number of students taking the UWR that semester, the standard is adjusted. In the last few years, educational reform effort in Kentucky has resulted in a new funding formula for institutions of higher education. Subsequently, this University has been concentrating very hard on increasing its recruitment and retention rate. It is possible that this factor is consciously or unconsciously affecting the interpretation and application of the scoring rubric by the scorers.

One of the seven readers who scored the essays for the study in year 2000 had read and scored the UWR in 1989. The seven readers used for this study had read and scored the UWR in 1997. It is possible but not probable that any of the readers scored any of the 50 essays in each group that were randomly selected for this study. Even if the readers did score any of the essays included in the study it is almost impossible that they could have remembered the essay(s) or the

score(s) assigned eleven or three years ago. Therefore, the probability of any contamination in scoring can safely be ruled out.

Results of this study viewed in the light of the results obtained in last year's study, which showed that the writing skills of pre-and post KERA students, as measured by the UWR, were not statistically significant, deserve re-examination of last year's results. Given the fact that this study is showing a slight inflation in scoring over the years, it is possible that had the readers of the 1980s been scoring the essays written by the students in the late1990s, a different result might have been found in last year's study.

Conclusions

A mean score of 8.18 assigned to a group of 50 UWR essays in 1989 and a mean score of 8.42 to the same group of essays by UWR readers in year 2000 does not show a statistically significant difference. Neither does a mean score of 8.16 assigned to the essays in 1997 and a mean score of 8.20 assigned to the same essays in 2000. However, there is a slight increase in the mean and this increase is greater between 1989-2000 than 1997-2000. In other words, the variance is larger over an eleven-year period compared to the variance over a three-year.

A correlation coefficient of 0.49 between 1989-2000 scorers is lower than the correlation coefficient of 0.78 between 1997-2000 scorers. Both are positive and significant. The inter-rater reliability is good but is growing weaker with passage of time.

The results of this study raise doubts about the results obtained in last year's study that compared the writing skills of pre-KERA students with the post-KERA students. No statistically significant difference was found between the means. The mean of the pre-KERA group was slightly higher compared to that of the post-KERA group, 8.18 and 8.16. A slightly lower mean obtained of the post-KERA group combined with a slight inflation in scoring needs to be noted and kept in mind. Further research is needed to come to a conclusion in that regard.

References

Department of Education. *2000. The KIRIS Accountability Cycle 3 Technical Report.*
Eastern Kentucky University. *The 1999-2001 Undergraduate Catalog.*
Legislative Research Commission. *The Kentucky Education Reform Act.* 1994 Edition.
National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, D. C.: U. S. Government Printing Office.
Sultana, Qaisar & Kay, Lisa. A Comparison of Pre-and Post KERA Students Writing Skills. Proceeding of the 29th Annual Conference of Mid-South Educational Association. 2000.

Table 1. UWR Policy at Eastern Kentucky University

To ensure that graduates of Eastern possess important communication skills, the faculty and Board of Regents have approved a University Writing Requirement (UWR). Except as noted below, students seeking baccalaureate degrees from Eastern, including transfer students, must successfully complete an essay exam in English.

Baccalaureate degree students must take the exam the first semester of enrollment after completing the 60th credit hour. Transfer students who transfer 60 credit hours or more must take the exam the first semester or enrollment.

Students who fail the first attempt may retake the exam under the following conditions:

A. prior to the next enrollment; they must file with their advisor a remediation plan;

B. they may not enroll for more than 12 hours in any fall or spring semester until the exam requirement is satisfied; and

C. they may not enroll after earning 100 hours until the exam requirement is satisfied.

Students failing to register for and take the UWR in the semester after they complete 60 credit hours will be subject to the enrollment limitations noted above in B. and C. Students with previously earned baccalaureate degrees need not write the UWR.

Table 2. UWR Scoring Rubric

The 7 - 6 paper responds to the prompt clearly and appropriately with sophisticated ideas;
- is well-organized, with effective transitions between ideas;
- develops key ideas coherently and effectively with details that have substance, specificity, or illustrative quality;
- has varied sentence structure and employs language that is vivid, precise, and fluent; demonstrates mastery of sentence structures, grammar, and other mechanics.
- There may be an occasional lapse from Standard English.

The 5 aper responds to the prompt with substantial ideas;
- is clearly organized;
- supports each idea with appropriate details.
- Its sentence structure, language choices, and use of Standard English may be flawed but are generally more than adequate.

The 4 paper covers the prompt with adequate organization,
- provides meaningful support for each idea, though perhaps some ideas are supported more effectively than others;
- has little variety in sentence structures although such structures may be adequate.
- Some errors in use of Standard English may occur.

The 3 paper does not respond adequately to the prompt either because it
- does not fully address the prompt,
- lacks coherent organization,
- lacks appropriate meaningful supportive detail,
- simply lists details without integrating them into the line of reasoning, or
- uses sentence structure and word choice that are inappropriately simple and repetitive.
- Spelling, punctuation, grammar, or syntax often fails to conform to Standard English

The 2 paper is deficient either because it
- shows little or no understanding of the prompt and/or the subject matter.
- shows little or no organization, or
- demonstrates little or no development of ideas.
- Inadequate sentence structure, word choices, or usage interferes with communication.

The 1 paper fails to communicate coherently either because it
- does not respond to the prompt,
- does not demonstrate organization or development of ideas.
- Fundamentally deficient sentence structures, word choices, and use of Standard English seriously interfere with communication.
- It is too brief to be an adequate sample of writing skills.

Table 3. The Writing Prompt

Spectator sports have assumed a prominent position today. Such sports receive a great deal of television, radio, newspaper, and magazine coverage; thousands of people flock to stadiums to watch the events.

What is the impact of spectator sports on your culture?

In a well-organized essay, state your view, provide reasons for your opinion, and support it with specific examples from your observation, experience, and academic studies.

****

Table 4- Statistical Procedures and Results

$\bar{X}$ of 1989 scores = 8.18, $s = 1.83$

$\bar{X}$ of scores assigned in 2000 = 8.42, $s = 1.40$

Pearson correlation coefficient = 0.49

$\bar{X}$ of 1997 scores = 8.16, $s = 1.98$

$\bar{X}$ of scores assigned in 2000 = 8.20, $s = 1.71$

Pearson correlation coefficient = 0.78

N= 50, Significant at .05 level on a two tail test 0.2